ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Physica A 351 (2005) 448–460

PHYSICA A

www.elsevier.com/locate/physa

Sum rules for free energy and frequency distribution of DNA dinucleotides

L. Frappat[a,*,1], A. Sciarrino[b]

[a]*Laboratoire d'Annecy-le-Vieux de Physique Théorique LAPTH, CNRS, UMR 5108, and Université de Savoie, BP 110, F-74941 Annecy-le-Vieux Cedex, France*
[b]*Dipartimento di Scienze Fisiche, Università di Napoli "Federico II" and I.N.F.N., Sezione di Napoli, Complesso Universitario di Monte S. Angelo, Via Cintia, I-80126 Naples, Italy*

Abstract

The large discrepancy between the values of the free energy for DNA dinucleotides (or dimers) measured by different teams has raised a debate, yet unsettled. Here the free energy is fitted by a three parameter empiric formula derived in the framework of the crystal basis model of genetic code. Approximate sum rules are derived and compared satisfactorily with the data. On the basis of theoretical and phenomenological arguments, a relation between the correlation functions of dimer distribution and the free energy is assumed. From consistency conditions, sum rules are derived. A check of these conditions with different samples of experimental data is performed, allowing us to argue on the self-consistency and the reliability of the different sets of experimental data.

PACS: 87.10.+e; 02.10.−v

Keywords: DNA dinucleotides; Free energy; Sum rules

*Corresponding author.
E-mail address: frappat@lapp.in2p3.fr (L. Frappat).
[1]Member of Institut Universitaire de France.

0378-4371/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2005.01.050

1. Introduction

The importance of computation of free energy ΔG^0 and enthalpy ΔH^0 for DNA dinucleotides or dimers was recognized in the eighties by many authors and several experimental measures have been performed. The experimental values however range in an unacceptably wide range. The basic model for computation of thermodynamical quantities for DNA and RNA is the nearest-neighbour model (NN), proposed in the sixties, which assumes that the thermodynamics is mainly governed by the interaction of two nearest nucleotides. However, a crucial point is the procedure to assign the corresponding value to the dimer, from the measured values of the thermodynamical quantities for different sequence combinations of oligomers and polymers. It was already pointed out in 1970 [1] that constraints that reduces the number of independent quantities in the NN model have to be taken into account. A turning point has been the introduction in Ref. [2] of a fictitious nucleotide mimicking the effect of the beginning and end of the sequences. Taking into account this idea, a detailed discussion of the number of constraints, both in single and double strand DNA, has been performed given in Ref. [3]. Moreover, the computation of the free energy of nucleotides sequences depends also on the helix initiation. A few years ago, SantaLucia [4] has performed an accurate analysis and comparison of the data from seven laboratories (see Table 8 taken from Ref. [4], where we have replaced the original values of the column Benight [5] with the more recent ones [6]), with the aim of presenting all the data in the same format. He reached the conclusion that six of the studies were actually in agreement and provided explanations for the discrepancies, even if the self-consistency of the data and the consistency between different data sets still remain debatable, and indeed urged for further experimental determination. In an attempt to settle, by the thermodynamics arguments, the controversy, Miramontes and Cocho [7] have analysed quite recently the same set of data by assuming a relation between the correlation function of the dimers and their free energy, reaching the conclusion that the most reliable set of values is just the one which was excluded by SantaLucia. Indeed, in Ref. [7] a linear relation between the correlation function for the dimer and the corresponding free energy was postulated, which allowed these authors to determine which set of experimental data was in better agreement with the postulated relation. A shortcoming of this analysis is that the sum of the free energies for strong dimers does not satisfy an identity derived from the postulated equation.The main purpose of this work is to come back to this controversial question. It is not our aim to analyse the methods used to derive the NN parameters, which is a complex and interesting subject widely discussed in the literature, but to look for reasonable criteria to evaluate the consistency of the data and, consequently, their reliability. First, we propose a theoretical formula to compute the free energy, from which sum rules are derived and compared with the values of experimental data. Second, we motivate the assumption of a relation between the correlation function and the free energy, different from the one assumed in Ref. [7], which satisfies trivial identities required by the definition of the correlation functions. We make several consistency checks and we try to determine the reliability of the

experimental values, comparing with the calculated values of the correlation matrix in Ref. [7].

2. Fit for the free energy

Let us recall that a mathematical framework in which the codons appear as composite states of nucleotides was proposed [8]. In order to make this paper self-contained, we briefly recall in the Appendix the main properties of $\mathscr{U}_{q\to 0}(sl(2))$, referring for more details to [8] or, for a more rigorous and mathematical discussion, to the original paper [9]. The four nucleotides are assigned to the fundamental irreducible representation of the quantum group $\mathscr{U}_q(sl(2)_H \oplus sl(2)_V)$ in the limit $q \to 0$—the indices H and V distinguish the two $sl(2)$—as follows:

$$\begin{array}{ccc} & sl(2)_H & \\ C \to \psi_{(1/2,1/2)_H,(1/2,1/2)_V} & \longleftrightarrow & U \to \psi_{(1/2,-1/2)_H,(1/2,1/2)_V} \\ sl(2)_V \updownarrow & & \updownarrow sl(2)_V \\ G \to \psi_{(1/2,1/2)_H,(1/2,-1/2)_V} & \longleftrightarrow & A \to \psi_{(1/2,-1/2)_H,(1/2,-1/2)_V}\,. \\ & sl(2)_H & \end{array} \tag{1}$$

A sequence of N nucleotides is then described by a pure state in the N-fold tensor product of the fundamental representation. In particular, dimers or dinucleotides are obtained as the two-fold tensor product, the labels which specify the irreducible representation to which they belong are given in Table 1 (see Appendix for details about the computation of the entries of this table). In Ref. [8] we have fitted old experimental data of the free energy ΔG^0_{37} (for simplicity we will omit the temperature label in the following) for RNA dinucleotides with a four parameter formula built up with the generators of $\mathscr{U}_{q\to 0}(sl_H(2) \oplus sl_V(2))$ and in Ref. [10] the more recent data of [11] have been fitted with the following two parameter formula:

$$\Delta G^0 = \alpha + \beta(C_H + C_V)J_{3H}\,, \tag{2}$$

where J_{3X} ($X = H$ or V) stands for the diagonalized $sl(2)_X$ generator and C_X is the Casimir operator of $\mathscr{U}_{q\to 0}(sl(2)_X)$ for the considered dimer ij. Let us recall that the

Table 1
Dimer representation content

Dimer	J_H	J_V	J_{3H}	J_{3V}	Dimer	J_H	J_V	J_{3H}	J_{3V}
CC	1	1	1	1	GC	1	1	1	0
CT	0	1	0	1	GT	0	1	0	0
CG	1	0	1	0	GG	1	1	1	−1
CA	0	0	0	0	GA	0	1	0	−1
TC	1	1	0	1	AC	1	1	0	0
TT	1	1	−1	1	AT	1	1	−1	0
TG	1	0	0	0	AG	1	1	0	−1
TA	1	0	−1	0	AA	1	1	−1	−1

Casimir operator eigenvalue in the J-representation is $J(J+1)$ (see formulae (40) and (41) of Appendix). In order not to overload the notation, here and in the following, we will not explicitly write the labels of the dimer, if not necessary to identify a specific dimer.

Here we propose for the DNA dinucleotides a three parameter formula, which is a generalisation of Eq. (2):

$$\Delta G^0 = \alpha_0 + \alpha_1 J_{3H} + \alpha_2 (J_{3V})^2 (2J_{3H} + 1)\,. \tag{3}$$

Using Table 1, this equation leads to theoretical values of the dimer-free energies ΔG^0 in terms of the parameters α_0, α_1, α_2, which are reported in Table 2.

A best-fit procedure allows one to evaluate these parameters. Indeed, one considers the square mean deviation between the theoretical and experimental dimer free energies ΔG^0 given by

$$s^2 = \frac{1}{N} \sum_{dimers} (\Delta G^0_{th} - \Delta G^0_{exp})^2\,, \tag{4}$$

where N is the number of points (here $N = 10$), the values of ΔG^0_{th} are given by Table 2 and the ΔG^0_{exp} correspond to a given set of experimental data. Minimizing the function s^2 with respect to the parameters α_i leads to the following expressions of these parameters:

$$\alpha_0 = \frac{1}{116}(14N_1 + 4N_2 - 6N_3), \quad \alpha_1 = \frac{1}{116}(4N_1 + 26N_2 - 10N_3),$$
$$\alpha_2 = \frac{1}{116}(-6N_1 - 10N_2 + 15N_3) \tag{5}$$

where (we specify by a couple of indices the free energy of a dinucleotide)

$$N_1 = \Delta G^0_{GG} + \Delta G^0_{CG} + \Delta G^0_{GC} + \Delta G^0_{CT} + \Delta G^0_{GA} + \Delta G^0_{GT} + \Delta G^0_{CA} + \Delta G^0_{TA} + \Delta G^0_{AT} + \Delta G^0_{AA},$$

$$N_2 = \Delta G^0_{GG} + \Delta G^0_{GC} + \Delta G^0_{CG} - \Delta G^0_{AA} - \Delta G^0_{AT} - \Delta G^0_{TA},$$
$$N_3 = 3\Delta G^0_{GG} + \Delta G^0_{CT} + \Delta G^0_{GA} - \Delta G^0_{AA}. \tag{6}$$

In Eq. (6) the dimer-free energies correspond to the given set of experimental values ΔG^0_{exp}. Hence we get in Table 3 for the different studies, see Table 8, the best-fit values of the parameters α_0, α_1, α_2. The last two rows correspond to the square mean

Table 2
Theoretical values of the dimer free energies ΔG^0

AA/TT	$\alpha_0 - \alpha_1 - \alpha_2$	CT/GA	$\alpha_0 + \alpha_2$
AT/TA	$\alpha_0 - \alpha_1$	GA/CT	$\alpha_0 + \alpha_2$
TA/AT	$\alpha_0 - \alpha_1$	CG/GC	$\alpha_0 + \alpha_1$
CA/GT	α_0	GC/CG	$\alpha_0 + \alpha_1$
GT/CA	α_0	GG/CC	$\alpha_0 + \alpha_1 + 3\alpha_2$

Table 3
Best-fit values of the α parameters for different studies

	Gotoh [12]	Vologodskii [13]	Breslauer [14]	Delcourt [15]	SantaLucia [16]	Sugimoto [17]	Unified [18]	Benight [6]
α_0	0.98	1.37	1.89	1.24	1.53	1.71	1.47	1.35
α_1	0.70	0.60	0.99	0.61	0.66	0.81	0.73	0.54
$-\alpha_2$	0.14	0.12	0.18	0.09	0.15	0.16	0.14	0.03
s^2	0.0015	0.0011	0.1577	0.0014	0.0114	0.0199	0.0070	0.0069
χ^2	0.0243	0.0099	1.0001	0.0167	0.0753	0.0992	0.0821	0.0590

deviation s^2 and to $\chi^2 = \sum(\Delta G^0_{exp} - \Delta G^0_{th})^2/\Delta G^0_{th}$. Evaluation of the incomplete Gamma function, which is an estimate of the goodness-of-fit, shows that the fit is good with a confidence level greater than 95%. Table 9 gives the fitted absolute values for dimer-free energy parameters ΔG^0 corresponding to the different samples, using formula (3) and the best-fit values of the parameters α_0, α_1, α_2 for each sample given by Table 3. From an inspection of the values of s^2 and χ^2, one sees that Eq. (3) is well fitted by the different sets of experimental data, except by the ones from Breslauer.

3. Sum rules

We derive from Eq. (3), a set of identities and sum rules. Let us first of all point out that the following sum rules are, of course, expected to be only approximately satisfied, as they are derived by empirical fitting formulae, not by a rigorous mathematical derivation from a theory. First, it is clear that

$$\Delta G^0_{ij} = \Delta G^0_{ji} \quad \text{and} \quad \sum_{j=A,C,G,T} \Delta G^0_{ij} = \sum_{j=A,C,G,T} \Delta G^0_{ji} \,. \tag{7}$$

In particular, we get

$$\sum_{j=A,C,G,T} \Delta G^0_{Cj} = \sum_{j=A,C,G,T} \Delta G^0_{Gj} = 4\alpha_0 + 2\alpha_1 + 4\alpha_2 \,, \tag{8}$$

$$\sum_{j=A,C,G,T} \Delta G^0_{Aj} = \sum_{j=A,C,G,T} \Delta G^0_{Tj} = 4\alpha_0 - 2\alpha_1 \,, \tag{9}$$

$$\sum_{i,j=A,C,G,T} \Delta G^0_{ij} = 16\,\alpha_0 + 8\,\alpha_2 \,. \tag{10}$$

In Table 4 we report the experimental values computed using the values of Table 8. Note that in Ref. [7] the existence of the sum rules Eqs. (8) and (9) was already remarked, but the two equations should have the same values, which is actually not the case.

Table 4
Experimental values of the sums of free energies [see Eq. (7)]

	Gotoh	Vologodskii	Breslauer	Delcourt	SantaLucia	Sugimoto	Unified	Benight
$\sum_i \Delta G^0_{Ci}$	4.72	6.16	9.18	5.78	6.72	8.10	6.74	6.54
$\sum_i \Delta G^0_{Gi}$	4.77	6.20	8.11	5.80	6.94	7.40	6.82	6.26
$\sum_i \Delta G^0_{Ti}$	2.55	4.27	5.63	3.68	5.08	5.30	4.33	4.45
$\sum_i \Delta G^0_{Ai}$	2.51	4.21	5.33	3.74	4.51	5.10	4.60	4.27

Table 5
Sum rules for free energies [see Eqs. (16)–(17)]

	Gotoh	Vologodskii	Breslauer	Delcourt	SantaLucia	Sugimoto	Unified	Benight
S_1	−0.07	0.08	2.19	0.10	−0.09	0.50	0.09	−0.32
S_2	−0.22	0.24	3.30	−0.14	0.34	0.60	0.52	0.36

Due to the complementarity rule, one has

$$\sum_{i=A,C,G,T} \Delta G^0_{Ci} = \sum_{i=A,C,G,T} \Delta G^0_{iG} \quad \text{and} \quad \sum_{i=A,C,G,T} \Delta G^0_{Gi} = \sum_{i=A,C,G,T} \Delta G^0_{iC} \,, \tag{11}$$

$$\sum_{i=A,C,G,T} \Delta G^0_{Ai} = \sum_{i=A,C,G,T} \Delta G^0_{iU} \quad \text{and} \quad \sum_{i=A,C,G,T} \Delta G^0_{Ui} = \sum_{i=A,C,G,T} \Delta G^0_{iA} \,. \tag{12}$$

Now we derive also news sum rules

$$\Delta G^0_{CG} + \Delta G^0_{TA} = 2\Delta G^0_{TG} = 2\Delta G^0_{AC} \,, \tag{13}$$

$$\Delta G^0_{CC} + \Delta G^0_{TT} = 2\,\Delta G^0_{TC} = 2\,\Delta G^0_{GA} \,, \tag{14}$$

$$\Delta G^0_{CC} + \Delta G^0_{AA} = 2\,\Delta G^0_{TC} = 2\,\Delta G^0_{AG} \,. \tag{15}$$

We report in Table 5 a comparison with the experimental data, making an average of the different experimental values, theoretically equal due to Eq. (3), i.e.,

$$S_1 = \Delta G^0_{CG} + \Delta G^0_{TA} + \Delta G^0_{GC} + \Delta G^0_{AT} - \Delta G^0_{TG} - \Delta G^0_{GT} - \Delta G^0_{AC} - \Delta G^0_{CA} = 0 \,, \tag{16}$$

$$S_2 = \Delta G^0_{CC} + \Delta G^0_{TT} + \Delta G^0_{GG} + \Delta G^0_{AA} - \Delta G^0_{CT} - \Delta G^0_{TC} - \Delta G^0_{AG} - \Delta G^0_{GA} = 0 \,. \tag{17}$$

As it can be seen the sum rules are reasonably well satisfied, except for the data of Breslauer. However we cannot make any statement on the reliability of the different experimental data on the basis of the accuracy by which they fit our empirical formula Eq. (3).

4. Dinucleotide distribution

In order for our analysis to settle on more theoretical ground, we consider the dimer correlation function. In [7] the dimer distribution was characterized by the correlation function

$$\Gamma_{ij} = f_{ij} - f_i f_j \,, \tag{18}$$

where the labels i,j denote the nucleotides, $i,j \in \{A, C, G, T\}$, and f_i (f_{ij}) denote the frequency of the i nucleotide (ij dinucleotide). From Eq. (18), it follows that

$$\sum_{i=A,C,G,T} \Gamma_{ij} = \sum_{j=A,C,G,T} \Gamma_{ij} = 0 \,. \tag{19}$$

In [7] the following relation between Γ_{ij} and the free energy ΔG^0 was assumed:

$$\Gamma_{ij} = a + b\Delta G^0_{ij} \,. \tag{20}$$

where a and b are biological species dependent parameters. Inserting Eq. (3) into Eq. (19) one gets the identity

$$4a + b \sum_{j=A,C,G,T} \Delta G^0_{ij} = 0 \Rightarrow \sum_{j=A,C,G,T} \Delta G^0_{ij} = const. \text{ for all } i \,. \tag{21}$$

In Ref. [7], from the data reported in Table 8, except the last column which was not considered, the authors show that Eq. (21) was satisfied by the weak dimers only, i.e., with label $i \in \{A, T\}$. Let us remark: (i) that the statistical mechanics motivation which led the authors to postulate Eq. (20) holds for an isolated system, which is not the case for a dimer inserted in a DNA strand; (ii) the computed values of the correlation matrix, see Table 3 of [7], for the same biological species, show in many cases, a much larger variation than the corresponding variation of the free energy, changing the ij dimer; (iii) our empirical formula Eq. (3) predicts the dimers ij and ji to have the same free energy, which is approximately true (see Table 8), while on the contrary the correlation function Γ_{ij} is generally non symmetric. From the above remarks we assume the following relation between Γ_{ij} and ΔG^0_{ij}:

$$\Gamma_{ij} = a + b\left(\Delta G^0_{ij} - \frac{1}{4}\sum_{k=A,C,G,T} (\Delta G^0_{ki} + \Delta G^0_{jk})\right) + (1 - \delta_{ij})h_{ij} \,, \tag{22}$$

where h_{ij} are biological species-dependent real coefficients. The complementarity implies that the coefficients h_{ij} and $h_{\overline{ji}}$ are equal for two complementary dimers ij (from $5'$ to $3'$) and $\overline{ji}$ (from $3'$ to $5'$), so there is only 8 coefficients h_{ij}.

The corrective term in the free energy can be considered as a "penalty" due to the interaction of the nucleotides of the dimer with the two nearest neighbour nucleotides in the strand, assumed uniformly distributed.

Since the correlation coefficient Γ_{ij} has to satisfy the sum rule (19) by definition, one is led to the constraints ($\forall j$)

$$\begin{aligned} 0 &= 4a + b \sum_{i=A,C,G,T} (\Delta G^0_{ij} - \Delta G^0_{ji}) - \frac{b}{4} \sum_{k,i=A,C,G,T} \Delta G^0_{ki} + \sum_{i=A,C,G,T} (1-\delta_{ij}) h_{ij} \\ &= 4a + b \sum_{i=A,C,G,T} (\Delta G^0_{ji} - \Delta G^0_{ij}) - \frac{b}{4} \sum_{k,i=A,C,G,T} \Delta G^0_{ik} + \sum_{i=A,C,G,T} (1-\delta_{ij}) h_{ij} \,. \end{aligned} \tag{23}$$

Eqs. (7)–(10) imply for any pair (i,j) of nucleotides

$$2b(2\alpha_0 + \alpha_2) - 4a = \sum_{k=A,C,G,T} (1-\delta_{ik}) h_{ik} = \sum_{k=A,C,G,T} (1-\delta_{kj}) h_{kj} \,. \tag{24}$$

As Eq. (24) gives 4 independent relations, we are left with 4 parameters h_{ij}. We remark that in Eq. (22) only the following combinations of a, b and α_i parameters appear in the free energy term:

$$x = a - b\alpha_0 \quad \text{and} \quad y = b\alpha_2 \,. \tag{25}$$

We then deduce from the 4 constraints (24) the following relations among the coefficients h_{ij} (we choose h_{CA}, h_{CT}, h_{CG}, h_{AC}, h_{TC}, h_{GC}, h_{AT}, h_{AT})

$$h_{CG} + h_{GC} - h_{AT} - h_{TA} = 0 \,, \tag{26}$$

$$h_{TC} - h_{CT} + h_{GC} - h_{AT} = 0 \,, \tag{27}$$

$$h_{CA} - h_{AC} + h_{CT} - h_{TC} + h_{CG} - h_{GC} = 0 \,. \tag{28}$$

Using Eq. (22) we can replace the following equations by sum rules for the corresponding correlation coefficients:

$$\Gamma_{CG} + \Gamma_{GC} - \Gamma_{AT} - \Gamma_{TA} = -4y = 2(\Gamma_{AA} - \Gamma_{CC}) \,, \tag{29}$$

$$\Gamma_{CT} - \Gamma_{TC} + \Gamma_{CG} - \Gamma_{TA} = -2y = \Gamma_{AA} - \Gamma_{CC} \,, \tag{30}$$

$$\Gamma_{CA} - \Gamma_{AC} + \Gamma_{CT} - \Gamma_{TC} + \Gamma_{CG} - \Gamma_{GC} = 0 \,. \tag{31}$$

The above equations are well satisfied (within $<5\%$) by the experimental data, see Table 3 of [7]. Therefore we conclude that our parametrisation (22) for the correlation function is satisfactory and we can carry on our analysis.

Consider the following differences of the correlation coefficients: $\Gamma_{CT} - \Gamma_{TC}$, $\Gamma_{TT} - \Gamma_{CC}$ and $\Gamma_{AT} - \Gamma_{GC}$. Inserting the theoretical expression (22) of Γ_{ij}, one gets for each of the three differences:

$$\Gamma_{CT} - \Gamma_{TC} = Z_{CT-TC} b + h_{CT} - h_{TC} \,, \tag{32}$$

$$\Gamma_{TT} - \Gamma_{CC} = Z_{TT-CC} b + h_{TT} - h_{CC} \,, \tag{33}$$

$$\Gamma_{AT} - \Gamma_{GC} = Z_{AT-GC} b + h_{AT} - h_{GC} \,, \tag{34}$$

where the coefficients Z are functions of the free energies ΔG^0, computed from Eq. (22). Summing up the three above equations, one gets that the l.h.s. is vanishing, due to Eq. (19) and the equality of the correlation coefficients for complementary dimers, which implies, using Eq. (27), that the coefficients Z are related:

$$Z_{CT-TC} + Z_{TT-CC} + Z_{AT-GC} = 0 \,. \tag{35}$$

Let us emphasize that this relation is biological species independent, by virtue of Eq. (27), valid for each biological species, and by the complementarity rule for Γ_{ij}.

Note also that relation (35) is automatically satisfied when plugging the theoretical expressions of the free energies of the dimers (i.e., in terms of the parameters α_0, α_1 and α_2).

Analogously using Eq. (28) and the complementarity rule we get

$$Z_{CA-GT} + Z_{CT-GA} + Z_{CG-GC} = 0 \,. \tag{36}$$

Note that Eq. (29) is satisfied identically from the parametrization (22) and the constraint (26).

We report in Tables 6 and 7 the values of the coefficients Z and their sum, calculated with the experimental free energies given by the different authors (see Table 8). As it can be seen, most of the values of the sums are quite close to zero, except for Breslauer, SantaLucia and Sugimoto.

Table 6
Values of the coefficients Z of Eq. (35)

	Gotoh	Vologodskii	Breslauer	Delcourt	SantaLucia	Sugimoto	Unified	Benight
Z_{CT-TC}	−0.123	−0.115	0.133	0.060	−0.498	0.125	0.027	0.005
Z_{TT-CC}	0.318	0.220	0.492	0.160	0.268	0.375	0.318	0.080
Z_{AT-GC}	−0.285	−0.205	0.145	−0.180	−0.560	0	−0.155	0.015
sum	−0.090	−0.100	0.770	0.040	−0.790	0.500	0.190	0.100

Table 7
Values of the coefficients Z of Eq. (36)

	Gotoh	Vologodskii	Breslauer	Delcourt	SantaLucia	Sugimoto	Unified	Benight
Z_{CA-AC}	−0.013	0.025	1.013	−0.110	0.358	0.425	−0.077	0.405
Z_{CT-TC}	−0.123	−0.115	0.133	0.060	−0.498	0.125	0.027	0.005
Z_{CG-GC}	0.035	0.010	0.995	0.010	−0.300	0.850	−0.110	0.150
sum	−0.100	−0.080	2.140	−0.040	−0.440	1.400	−0.160	0.560

Table 8
Experimental absolute values for dimer free energy parameters ΔG^0 (in kcal/mol)

	Gotoh [12]	Vologodskii [13]	Breslauer [14]	Delcourt [15]	SantaLucia [16]	Sugimoto [17]	Unified [18]	Benight [6]
AA/TT	0.43	0.89	1.66	0.67	1.02	1.20	1.00	0.91
AT/TA	0.27	0.81	1.19	0.62	0.90	0.90	0.88	0.83
TA/AT	0.22	0.76	0.76	0.70	0.90	0.90	0.58	0.68
CA/GT	0.97	1.37	1.80	1.19	1.70	1.70	1.45	1.54
GT/CA	0.98	1.35	1.13	1.28	1.43	1.50	1.44	1.25
CT/GA	0.83	1.16	1.35	1.17	1.16	1.50	1.28	1.28
GA/CT	0.93	1.25	1.41	1.12	1.46	1.50	1.30	1.30
CG/GC	1.70	1.99	3.28	1.87	2.09	2.80	2.17	1.87
GC/CG	1.64	1.96	2.82	1.85	2.28	2.30	2.24	1.86
GG/CC	1.22	1.64	2.75	1.55	1.77	2.10	1.84	1.85

5. Conclusions

We have proposed a three-parameter formula to fit the free energy for the DNA dinucleotides and derived a set of sum rules. Let us emphasize that the sum rules have to be considered as approximate identities derived from empirical formulae. Let us also remark that, in the comparison between the experimental and the theoretical values computed from Eq. (2), for RNA and for DNA, a larger discrepancy is present between the GC and CG dimer for RNA structure than for DNA. This feature can be understood as an effect of the more relevant role played in the thermodynamics by the GC content in DNA than in RNA; e.g. an empirical formula depending on four parameters has been derived in Ref. [19], expressing the melting temperature of DNA as a function of its fractional GC content and of the concentration on Na^+ ions. We have compared the theoretical values with the experimental data of seven authors as well as their averaged value. The results of the fits reported in Tables 5 and 9, show in the average a satisfactory agreement, except for Breslauer. On the basis of the above comparison, we cannot make any statement on the reliability of the different experimental data. In order to support our analysis by general theoretical arguments, we postulate a relation between the free energy and the dimer correlation function Eq. (22), which has theoretical motivation from statistical mechanics as well as experimental motivation from the analysis of the computed correlation function. Our postulated equation is self-consistent as it satisfies the identity that the sum of correlation functions has to satisfy by definition. From consistency equations, we derive a set of sum rules for the correlation functions which are well verified by the computed values for several biological species. This analysis supports the validity of our relation Eq. (22), which allows us to perform biological independent consistency checks, which is remarkably verified by our theoretical formula. We have checked which set of experimental data satisfy the consistency relations. The result is that the data of Refs. [14,16] and [17] are not

Table 9
Fitted absolute values for dimer free energy parameters ΔG^0 (in kcal/mol)

	Gotoh	Vologodskii	Breslauer	Delcourt	SantaLucia	Sugimoto	Unified	Benight
AA/TT	0.46	0.92	1.13	0.75	1.08	1.11	0.93	0.85
AT/TA	0.30	0.79	0.93	0.65	0.91	0.93	0.78	0.81
TA/AT	0.30	0.79	0.93	0.65	0.91	0.93	0.78	0.81
CA/GT	1.02	1.40	1.94	1.26	1.57	1.75	1.51	1.36
GT/CA	1.02	1.40	1.94	1.26	1.57	1.75	1.51	1.36
CT/GA	0.85	1.27	1.73	1.16	1.40	1.57	1.35	1.33
GA/CT	0.85	1.27	1.73	1.16	1.40	1.57	1.35	1.33
CG/GC	1.73	2.01	2.94	1.88	2.24	2.57	2.25	1.90
GC/CG	1.73	2.01	2.94	1.88	2.24	2.57	2.25	1.90
GG/CC	1.25	1.61	2.34	1.57	1.73	2.03	1.78	1.81

consistent. Therefore we disagree with the conclusions of Ref. [7]. The results of our analysis are more close to the ones of Ref. [4].

Obviously the "sum rules", whose approximate validity allows one to reduce the number of independent parameters in the NN model, can be formulated without making any reference to the "crystal basis" and they might have been derived from an analysis of the experimental data. In the crystal model, they are a straightforward derivation of the simple expression of an empirical formula to fit the free energy. Indeed, as it has been discussed in some details in Ref. [10], the crystal basis model seems to provide a useful mathematical setting to formulate some properties of DNA and/or RNA, which may imply that some essential physico-chemical features have been indeed incorporated in the mathematical language. Let us remark that a dimer with $J_{3H} = 0$ ($J_{3V} = 0$) means that it is made by two not identical purines or pyrimidines or by a purine and a not complementary pyrimidine (respectively, by a purine and a pyrimidine).

Acknowledgements

We would like to thank the referee for constructive remarks and for drawing to our attention some relevant references.

Appendix A. Basic notions on crystal bases

In this appendix we briefly recall the main properties of the so-called deformed universal enveloping algebra of $sl(2)$, denoted $\mathscr{U}_q(sl(2))$, and its limit $\mathscr{U}_{q\to 0}(sl(2))$. The algebra $\mathscr{U}_q(sl(2))$ is defined as a suitable completion of the algebra of polynomes in the generators $\widetilde{J}_+$, $\widetilde{J}_-$ and $\widetilde{J}_3$ (in particular adding the exponential series), subject

to the following commutation relations:

$$[\widetilde{J}_+, \widetilde{J}_-] = [2\widetilde{J}_3]_{\mathrm{q}},$$
$$[\widetilde{J}_3, \widetilde{J}_\pm] = \pm\widetilde{J}_\pm, \tag{37}$$

where

$$[x]_q = \frac{q^x - q^{-x}}{q - q^{-1}} \ . \tag{38}$$

Moreover some suitable axioms have to be fulfilled, which endows $\mathscr{U}_{\mathrm{q}}(sl(2))$ with a Hopf algebra structure. Since we do not need these axioms here, we do not explicitly write them for sake of simplicity.

The vector spaces of the irreducible representations of this algebra are labelled, for q different of root of unity, by a non negative integer or half-integer number j and are of dimension $(2j+1)$, the basis vectors being denoted by ψ_{jm}, $-j \leqslant m \leqslant j$. In the limit $q \to 1$ one recovers the usual $sl(2)$. Strictly speaking, in the limit $q \to 0$ the generators are ill defined, but it is possible, see Ref. [9], to define new generators $J_\pm$, $J_3 (= \widetilde{J}_3)$, whose action on the vector basis of the representation space, still labelled by a non negative integer or half-integer number j and of dimension $(2j+1)$, is well defined:

$$J_3\psi_{jm} = m\psi_{jm}, \quad J_\pm\psi_{jm} = \psi_{j,m\pm1}, \quad J_\pm\psi_{j,\pm j} = 0 \ . \tag{39}$$

This special basis in the limit $q \to 0$ is called a crystal base. Note that the action of $J_\pm$ on ψ_{jm} is equal to $\psi_{j,m\pm1}$ (i.e., the coefficient is always 1), contrary to the $sl(2)$ or $\mathscr{U}_{\mathrm{q}}(sl(2))$ case where this coefficient is a complicated function of j and m.

It is possible also to define an operator C called Casimir operator [8], such that

$$C\psi_{jm} = j(j+1)\psi_{jm} \Longrightarrow [C, J_\pm] = [C, J_3] = 0 \ . \tag{40}$$

Its explicit expression is given by

$$C = (J_3)^2 + \frac{1}{2}\sum_{n\in\mathbb{Z}_+}\sum_{k=0}^{n}(J_-)^{n-k}(J_+)^n(J_-)^k \ . \tag{41}$$

Although this Casimir operator is written as an infinite series of powers of $J_\pm$, in any crystal base, only a finite number of terms gives a non-vanishing contribution, which leads to (40).

Notice that $\mathscr{U}_{\mathrm{q}\to0}(sl(2))$ is neither a deformed universal enveloping algebra nor a Hopf algebra. However, one can show [9] that the tensor product of two crystal bases labelled by j_1 and j_2 can be decomposed into a direct sum of crystal bases labelled, as in the case of the tensor product of two $sl(2)$ or of $\mathscr{U}_{\mathrm{q}\to0}(sl(2))$ irreducible representations, by an integer or half-integer number j such that

$$|j_1 - j_2| \leqslant j \leqslant j_1 + j_2 \ . \tag{42}$$

The new peculiar and crucial feature, which is the key point in the model proposed in Ref. [8], is that now the basis vectors of the j-space are *pure states*, that is they are the product of a state belonging to the j_1-space and of a state belonging to the j_2-space,

while in the case of $sl(2)$ or of $\mathscr{U}_{q\to 0}(sl(2))$ they are linear combinations with coefficients called, respectively, Clebsch-Gordan coefficients or q-Clebsch-Gordan coefficients. As an example, we obtain for the dimer CT and TC, from Eq. (1) and from the rules to perform the tensor product (see [9,8]):

$$C \otimes T \equiv \psi_{(1/2,1/2)_H,(1/2,1/2)_V} \otimes \psi_{(1/2,-1/2)_H,(1/2,1/2)_V} = \psi_{(0,0)_H,(1,0)_V} \equiv CT\,,$$
$$T \otimes C \equiv \psi_{(1/2,-1/2)_H,(1/2,1/2)_V} \otimes \psi_{(1/2,1/2)_H,(1/2,1/2)_V} = \psi_{(1,0)_H,(1,0)_V} \equiv TC\,. \qquad (43)$$

References

[1] D.M. Gray, I. Tinoco, Biopolymers 9 (1970) 223–244.
[2] R.F. Goldstein, A.S. Benight, Biopolymers 32 (1992) 1679–1693.
[3] D.M. Gray, Biopolymers 42 (1997) 783–793.
[4] J. SantaLucia, Proc. Natl. Acad. Sci. USA 95 (1998) 1460–1465.
[5] M.J. Doktycz, R.F. Goldstein, T.M. Paner, F.J. Gallo, A.S. Benight, Biopolymers 32 (1992) 849–864.
[6] R. Owczarzy, P.M. Vallone, R. F Goldstein, A.S. Benight, Biopolymers 52 (1999) 29–56.
[7] P. Miramontes, G. Cocho, Physica A 321 (2003) 577–586.
[8] L. Frappat, A. Sciarrino, P. Sorba, Phys. Lett. A 250 (1998) 214–221.
[9] M. Kashiwara, Commun. Math. Phys. 133 (1990) 249–260.
[10] L. Frappat, A. Sciarrino, P. Sorba, J. Biol. Phys. 27 (2001) 1–34.
[11] D.H. Mathews, J. Sabina, M. Zucker, D.H. Turner, J. Mol. Biol. 288 (1999) 911–940.
[12] O. Gotoh, Y. Tagashira, Biopolymers 20 (1981) 1033–1042.
[13] A.V. Vologodskii, B.R. Amirikyan, Y.L. Lyubchenko, M.D. Frank-Kamenetskii, J. Biomol. Struct. Dyn. 2 (1984) 131–148.
[14] K.J. Breslauer, R. Frank, H. Blocker, L.A. Marky, Proc. Natl. Acad. Sci. USA 83 (1986) 9373–9377.
[15] S.G. Delcourt, R.D. Blake, J. Biol. Chem. 266 (1991) 15160–15169.
[16] J. SantaLucia, H. Allawi, P.A. Seneviratne, Biochemistry 35 (1996) 3555–3562.
[17] N. Sugimoto, S. Nakano, S. Yonemaya, K. Honda, Nucleic Acids Res. 24 (1996) 4501–4505.
[18] H.T. Allawi, J. SantaLucia, Biochemistry 36 (1997) 10581–10594.
[19] M.D. Frank-Kamenetskii, Biopolymers 10 (1971) 2623–2624.